Soy Formula and Epigenetic Modifications: Analysis of Vaginal Epithelial Cells from Infant Girls in the IFED Study

Sophia Harlid, Margaret Adgent, Wendy N. Jefferson, Vijayalakshmi Panduri, David M. Umbach, Zongli Xu, Virginia A. Stallings, Carmen J. Williams, Walter J. Rogan, and Jack A. Taylor

Soy Formula and Epigenetic Modifications: Analysis of Vaginal Epithelial Cells from Infant Girls in the IFED Study

Sophia Harlid[1,2], Margaret Adgent[2]*, Wendy N. Jefferson[3]*, Vijayalakshmi Panduri[1], David M. Umbach[4], Zongli Xu[2], Virginia A. Stallings[5], Carmen J. Williams[3], Walter J. Rogan[2], and Jack A. Taylor[1,2]

Author affiliations:

[1] Epigenetics & Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, North Carolina, USA

[2]Epidemiology Branch, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, North Carolina, USA

[3]Reproductive & Developmental Biology Laboratory, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, North Carolina, USA

[4]Biostatistics & Computational Biology Branch, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, North Carolina, USA

[5]Division of Gastroenterology, Hepatology and Nutrition, Children's Hospital of Philadelphia, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA

* These authors contributed equally to this work

Correspondence to: Jack A. Taylor, MD, PhD, National Institute of Environmental Health Sciences, MD A3-05, 111 T.W. Alexander Drive, PO Box 12233, Research Triangle Park, NC 27709. E-mail: taylor@niehs.nih.gov. Phone: (919) 541-4631, Fax : (919) 541-2511

Running Title: Soy formula exposure and epigenetic modifications

Competing Interests: The authors declare no competing interests

ABSTRACT

Background: Early life exposure to estrogenic compounds affects the development of the reproductive system in rodent models and humans. Soy products, which contain phytoestrogens such as genistein, are one source of exposure in infants fed soy formula and result in high serum concentrations.

Objectives: To determine if soy exposure is associated with differential DNA methylation in vaginal cells from soy-fed infant girls.

Methods: Using the Illumina HumanMethylation450 BeadChip we evaluated epigenome-wide DNA methylation in vaginal cells from four soy-formula-fed and six cow-formula-fed girls from the Infant Feeding and Early Development (IFED) study. Using pyrosequencing we followed up the two most differentially methylated sites in 214 vaginal cell samples serially collected between birth and nine months of age from 50 girls (28 soy-formula-fed and 22 cow-formula-fed). With a mouse model, we examined the effect of neonatal exposure to genistein on gene specific mRNA levels in vaginal tissue.

Results: The epigenome-wide scan suggested differences in methylation between soy-formula-fed and cow-formula-fed infants at three CpGs in the gene proline rich 5 like (*PRR5L*) (p<10^4). Pyrosequencing of the two feeding groups found that methylation levels progressively diverged with age, with pointwise differences becoming statistically significant after 126 days. Genistein exposed mice showed a 50% decrease in vaginal *Prr5l* mRNA levels compared to controls.

Conclusions: Girls fed soy formula have altered DNA methylation in vaginal cell DNA which may be associated with decreased expression of an estrogen-responsive gene.

INTRODUCTION

Soy formula use during infancy has been associated with altered age at menarche (Adgent et al. 2012; D'Aloisio et al. 2013) and identified as a risk factor for uterine fibroids (D'Aloisio et al. 2012) and endometriosis (Upson et al. 2015). These associations might be attributed to exposure to the phytoestrogen genistein, which is present in high amounts in soy formula and binds to the estrogen receptor (ER) with highest affinity for ERβ (Kuiper et al. 1998). Neonatal rodents exposed to genistein show alterations in their estrous cycle, behavior, ovarian function, and reproductive tract (reviewed in Jefferson et al. 2012), as well as increased risk of tumors and possibly autoimmune diseases (Klein et al. 2002; Newbold et al. 2001). Short term dietary exposure to genistein also alters DNA methylation patterns in mice (Day et al. 2002).

In the US, about 12% of formula-fed infants are given soy formula (Rossen et al. 2015). Although soy formulas are considered to be safe, infants fed soy formula have blood concentrations of genistein that overlap with those showing biologic effects in rodents (Cao et al. 2009; McCarver et al. 2011), raising concern about possible latent or subclinical effects. As yet, few data are available for examining this hypothesis. A prospective observational study of soy formula, cow milk formula and breast-fed infants in the central United States found no difference in mental development status at one year of age between infants fed soy formula or cow milk formula (Andres et al. 2012). In a later follow-up study, reproductive organ size was compared between children who had been fed different formulas or breast milk; again no statistically significant differences were found (Andres et al. 2015). However, reproductive organ size measurements may not be sensitive markers of long term estrogen effects. A pilot study with a focus on the reproductive tract (Bernbaum et al. 2008), and preliminary data from the subsequent

longitudinal study (Adgent et al. 2014); found that infant girls fed soy formula had vaginal cytological changes consistent with estrogen exposure. We hypothesized that early life genistein exposure from soy formula might lead to later changes in the reproductive system through epigenetic alteration of reproductive tract tissues. Here we report on differences in DNA-methylation patterns in vaginal epithelial cells between infant girls being exclusively fed soy formula or cow milk formula during their first months of life.

METHODS

Human study sample

All participants were part of the Infant Feeding and Early Development (IFED) Study, a longitudinal cohort study designed to identify differences in estrogen-responsive outcomes in infants fed cow milk-based formula, soy-based formula, or breast milk. Between August 2010 and March 2014, IFED enrolled mothers in the 3rd trimester of pregnancy or within the first 72 hours post-partum and followed their infants prospectively. Feeding group was determined by the mother's intention to feed a simple, exclusive regimen from birth; mother-infant pairs were excluded if their feeding method changed during follow up. Eligible mothers had to speak English, be ≥ 18 years of age and have no history of gestational diabetes, thyroid dysfunction, or other endocrine disorders. Eligible infants were healthy singletons, born between 37-42 weeks of gestation with a birth weight between 2500-4500 grams. Mothers provided written informed consent for both themselves and their infants. The IFED study protocol was approved by the institutional review boards at the National Institute of Environmental Health Sciences and the Children's Hospital of Philadelphia, the clinical site of the study.

All infant girls were followed until nine months of age. Visits occurred every two weeks until one month of age and then every four weeks until end of follow-up. At each visit, infants had a physical examination and had biological samples collected, including a swab taken from the introitus of the vagina for cytological analysis. Methods related to swab collection and sample treatment have been previously described (Adgent et al. 2013). The study enrolled 410 mother-infant pairs, 397 of which contributed at least one swab sample. Two hundred eighty-three infant boys and girls completed the full course of the study (70 breast-milk-fed, 111 cow-formula-fed and 102 soy-formula-fed).

Our study used DNA extracted from cytology swabs collected after December 17, 2012. At this time, recruitment and follow-up of breast-fed infants had concluded making those samples (all collected prior to December 17, 2012) unavailable for our analysis. The IFED study also enrolled boys who had swabs taken from the urethral meatus. DNA yields from the urethral samples were poor, so we excluded boys from the methylation study. Therefore, our analyses only included samples from cow- and soy-formula-fed girls.

Human sample processing

DNA was extracted from polyester swabs stored in SurePath Preservative (BD Diagnostics, Durham NC) using the QIAamp DNA blood mini kit (Qiagen). After extraction, 237 out of 464 available samples yielded sufficient DNA for further analysis. All extracted DNA was bisulfite converted using the EZ DNA methylation kit (Zymo Research) according to the manufacturer's protocol. The 20 swabs yielding most DNA were used for epigenome-wide DNA methylation analysis with Infinium HumanMethylation450 BeadChip arrays (Illumina Inc.), including two

samples run in duplicate. Input DNA ranged from 200-500 ng which is less than the 500 ng recommended by the manufacturer. Additional details are provided in Supplemental Materials.

The remaining 217 samples, from 52 girls, were used for pyrosequencing as described in the Supplemental Material and Table S1. During follow up two of the 52 infants changed feeding method and were therefore excluded, yielding a final sample size of 214 samples from 50 infant girls (155 from 28 soy-formula-fed and 59 from 22 cow-formula-fed). See Figure S1 for a complete description of sample selection. In order to provide direct comparison to 450K results we also pyrosequenced 10 samples that had been successfully analyzed on the 450K platform.

Methylation and gene expression

The relationship between DNA methylation and gene mRNA expression was examined using publically available Cancer Genome Atlas (TCGA) correlation analyses deposited at Broad GDAC Firehose (https://gdac.broadinstitute.org/).

Mouse model

Animals were handled according to NIH/NIEHS guidelines under approved animal care and use protocols. Timed pregnant CD-1 mice were obtained from the in-house breeding colony at NIEHS (Research Triangle Park, NC), housed in a temperature controlled environment (21-22º C) under a 12 h light: 12 h dark cycle, and fed NIH-31 diet. Sixty female pups were treated by subcutaneous injections (0.02 mL) on the day of birth (post-natal day 1) through post-natal day 5 with either genistein 50 mg/kg per day dissolved in corn oil (n=30) or with corn oil alone (controls; n=30) as described previously (Doerge et al. 2002). Treated mice were sacrificed on post-natal day 5 (20 exposed/20 control) and post-natal day 22 (10 exposed/10 control). Ten

additional female pups were sacrificed on the day of birth without any prior treatment. Mouse vaginal tissue was collected and stored at -80˚C. RNA extraction and expression analysis are described in the Supplemental Material.

Statistical analysis

Demographic characteristics were compared between feeding groups using chi-square and Fisher's exact tests.

Before analyzing our EWAS data we excluded 58,840 CpGs that either: mapped to multiple target regions, included SNPs or were located on the X or Y chromosome. We then tested the association between feeding regimen (i.e., soy or cow milk) and DNA methylation (β value) at each of the remaining 427,097 CpG sites from the genome-wide array using a robust linear regression model employed via the R package MASS (Venables and Ripley 2002) without correction for multiple comparisons.

To confirm our findings from the 450K analysis of human vaginal cell DNA we used pyrosequencing to measure % DNA methylation at two specific CpG sites in the gene proline rich 5 like (*PRR5L*). We created a single summary methylation (M) value (denoted $\overline{M}$) for each sample by averaging four individual M values, *i.e.*, those from duplicate runs for each of the two CpGs. We transformed $\overline{M}$ to $logit(\overline{M}) = \ln(\overline{M}/(1-\overline{M}))$ and used $logit(\overline{M})$ as the outcome in our statistical analysis. This transformation ensures that, upon back-transformation to the original scale, fitted M values will remain within the desired (0-1) range. We estimated age trajectories for $logit(\overline{M})$ for all subjects, and separately for each feeding group, using mixed-model techniques to account for variability both within and among subjects. We represented each mean trajectory as a natural cubic spline with three equally spaced knots using the square root of age

(days at sample collection) as the predictor; knot locations were common to both feeding groups. We accounted for variability among subjects by declaring all spline coefficients except the intercept to be random with an unstructured covariance matrix, in effect allowing each subject her own trajectory. Our model allowed the two feeding groups to have different within-subjects variances. Regarding the logit($\overline{\mathrm{M}}$) as having a normal distribution, we fit this mixed model using PROC GLIMMIX in SAS 9.3 (Cary, NC). We calculated back-transformed fitted means, variances, and confidence limits using Taylor series methods.

Gene expression was evaluated in the mouse model. Using newborn mice as a reference we quantified relative levels of *Prr5l* mRNA in mouse vaginal tissue. We compared the mean mRNA levels of *Prr5l* between the following mouse groups: newborn vs post-natal day 5 exposed, newborn vs post-natal day 5 controls, and post-natal day 5 exposed vs post-natal day 5 controls with the Mann-Whitney U test.

RESULTS

Epigenome-wide analysis

To identify differentially methylated loci in soy-formula-fed vs. cow-formula-fed infant girls, we used the Illumina HumanMethylation450 BeadChip, which tags 485,577 CpG sites across the genome. Perhaps reflecting sub-optimal input DNA, 9 samples and 1 duplicate failed QC metrics and were eliminated from subsequent analysis. Eleven vaginal cell DNA samples and one duplicate provided high quality methylation data for genomic analysis, corresponding to 4 soy-formula-fed and 6 cow-formula-fed subjects (1 cow-formula-fed subject was sampled once at 20

weeks and in duplicate at 24 weeks of age). The demographic characteristics of the mothers of these 10 infants were similar to those described in Table 1, with 6 of 10 mothers identified as black, 7 of 10 mothers 21-30 years of age, and 7 of 10 with a high school education or less. The samples from 4 soy-formula-fed girls were from younger ages (4 - 8 weeks of age) than the 6 samples from cow-formula-fed girls (20 – 32 weeks of age).

Five CpG sites had p-values $< 10^{-4}$ and also had effect sizes (difference in mean β-value between cow-formula-fed and soy-formula-fed) of >0.3 (Figure 1A). Cg13935577 is located in the promoter region of *BTBD11* on chromosome 12 and cg20103692 is in *MAS1L* on chromosome 6. The three remaining sites (cg00220721, cg22117805, and cg08943494), including the site with the smallest p value (cg22117805), are all located within 200 bp of one of the annotated transcriptional start sites of *PRR5L;* all three had β-values that were low in cow-formula-fed girls (<0.3) and high in soy-formula-fed girls (>0.5) (Figure 1B). Taken together, the small consistent p-values, large effect sizes, and close proximity of the three CpG sites to a *PRR5L* transcriptional start site led us to investigate these CpGs in a larger study that sought to assess both formula and age effects.

Pyrosequencing

The two *PRR5L* CpGs with closest proximity to each other (12 bp apart) were selected for pyrosequencing analysis (cg00220721 and cg22117805). Methylation levels at these two sites were highly correlated ($R^2 = 0.95$) on 450K array analysis and the average methylation at the 2 sites on 450K analysis was in turn highly correlated to the average methylation obtained by pyrosequencing ($R^2 = 0.96$, Figure 1C). We did not assess any additional CpGs at this or other loci.

To determine if we could replicate results from the epigenome-wide analysis in independent samples and to investigate age effects we used DNA from 214 vaginal cell DNA samples (distinct from those included in the epigenome-wide analysis) from 50 infant girls (Table 1). Of the 214 samples examined with pyrosequencing, our final analysis included 205 samples (147 soy-formula-fed and 58 cow-formula-fed) from 49 girls, excluding nine samples that failed in the duplicate sequencing runs for both sites. Mean methylation levels in the two feeding groups were similar and high, near 0.8, at the time of birth. Methylation levels decreased with increasing age ($p<0.001$) but soy-formula-fed girls maintained higher mean methylation levels compared to cow-formula-fed girls (Figure 2A). Although the overall trajectories for soy-formula and cow-formula-fed girls did not differ significantly ($p = 0.25$), methylation differences calculated at each time point (without correction for multiple comparisons) were statistically significant ($p<0.05$) beginning about 126 days after birth (Figure 2B).

PRR5L gene expression

In order to determine the relationship between methylation and mRNA expression of *PRR5L* we examined publically available Cancer Genome Atlas (TCGA) data through the Broad GDAC Firehose database. DNA methylation at cg0022072 is significantly negatively correlated with *PRR5L* mRNA level with Spearman correlations of -0.53 to -0.55 (Supplementary Table S2, Broad Institute TCGA Genome Data Analysis Center. 2015).

To determine if expression of *Prr5l* was affected by genistein exposure in vivo, we used a mouse model in which neonatal mice were exposed to genistein at doses that produced serum concentrations similar to those in human infants fed soy formula (Cao et al. 2009; Doerge et al. 2002). Relative to untreated newborn female mice, expression of *Prr5l* was higher in both

exposed and control mice at post-natal day 5, but genistein treated mice had significantly lower

expression (p<0.01) than control mice (Figure 3). By postnatal day 22, *Prr5l* expression had

fallen more than 10-fold and did not differ by prior treatment (data not shown). The vaginal

preparations from mouse that were used in this analysis include blood, connective tissue, and

other cell types that may be less estrogen-responsive and therefore underestimate the effect of

treatment on gene expression. These results from both human tumor tissue and mouse vaginal

tissue suggest that the effect of genistein on DNA methylation in humans would be consistent

with a corresponding decrease in expression of the *PRR5L* gene early in life.

DISCUSSION

Here we find that girls fed soy formula from birth exhibited higher DNA methylation at a

specific gene locus in their vaginal tissue compared to girls fed cow formula. Methylation

differences were observed at CpG sites flanking a *PRR5L* transcriptional start site. These

differences were confirmed by pyrosequencing of vaginal DNA samples that had been serially

collected over nine months following birth. Pyrosequencing revealed that DNA methylation at

PRR5L was high at birth and fell rapidly in the two months after birth. This drop in methylation

coincides with the infant's rapidly falling exposure to maternal estrogens (Bidlingmaier et al.

1973). Compared to cow-formula-fed girls, phytoestrogen-exposed soy-fed girls maintained

higher methylation levels over time. Correlation between DNA methylation at cg0022072 and

decreased gene expression was confirmed using deposited human data from the TCGA database.

The long term persistence of increased *PRR5L* DNA methylation in soy-fed infants and the

health consequences, if any, remain unknown.

Relatively little is known about *PRR5L*, which encodes a component of the mTORC2 complex. The PRR5L protein (also known as protor-2) suppresses mTORC2-mediated activation of protein kinase C delta, which promotes fibroblast migration (Gan et al. 2012). PRR5L also promotes apoptosis via activation of the pro-inflammatory cytokine, tumor necrosis alpha (TNFα) (Thedieck et al. 2007). When PRR5L dissociates from the mTORC2 complex in response to mTORC2 activation, it binds to tristetraprolin (TTP), an RNA-binding protein that mediates sequence-specific degradation of mRNAs (Carrick et al. 2004; Holmes et al. 2012). PRR5L interaction with TTP appears to be required for proper TTP localization to cytoplasmic granules where mRNA processing occurs; in the absence of PRR5L, TTP does not degrade its target mRNAs. One of the best-known mRNA substrates for TTP is TNFα: Mice lacking TTP have high levels of TNFα due to low mRNA turnover, leading to severe arthritis and cachexia (Taylor et al. 1996). This link between PRR5L, TTP, and TNFα is interesting in light of recent studies demonstrating the association of a SNP (rs4755450) in *PRR5L* with juvenile idiopathic arthritis and a microarray study showing that downregulation of *PRR5L* was associated with osteoarthritis in adults (Chiaroni-Clarke et al. 2014; Wang et al. 2015). These findings suggest that suppression of *PRR5L* expression might promote TNFα-associated autoimmune diseases, including common conditions such as asthma, rheumatoid arthritis, psoriasis, or inflammatory bowel disease.

Using a murine model we demonstrated that *Prr5l* is an estrogen-regulated gene, with genistein significantly suppressing *Prr5l* expression in vaginal tissue of neonatally-exposed mice. This regulation is perhaps not surprising given results of chromatin immunoprecipitation-sequencing showing the mouse *Prr5l* gene locus has three distinct sites of estrogen receptor α binding (Hewitt et al. 2012). Genistein affects mTOR signaling, leading to increased interest in its use in

cancer prevention and treatment (reviewed in Ahmad et al. 2013), but epigenetic modification of

PRR5L has not been previously reported.

Plant estrogens, along with other chemicals with estrogenic or other hormonal activity, are often

classified as "endocrine disrupters." Under the endocrine disruption hypothesis, exposure to

certain cosmetics, plasticizers, pesticide residues, dietary and other agents cause subtle

alterations in endocrine function, leading to subsequent adverse health effects. Evidence exists

for this hypothesis in whole animal model systems and in wildlife, with exposure during the

critical perinatal period producing the strongest effects (Diamanti-Kandarakis et al. 2009). But

even for the best studied agents, those that exhibit estrogenicity, human epidemiologic data of

health consequences have remained tenuous: e.g., the inference that endocrine disruptors may be

responsible for secular trends in sperm count, infertility and obesity (Bergman et al. 2013). In

part, the uncertainty in the human evidence reflects the substantial epidemiologic challenge of

connecting very low dose exposures during infancy and childhood to health effects that may only

become manifest years later in adolescents and adults. Soy formula feeding of infants provides a

useful window into this challenge because it involves early life exposure to much higher levels

of estrogenic compounds than would come from other putative endocrine disrupters (Behr et al.

2011; Riu et al. 2008).

Recent epidemiologic studies have shown that soy formula feeding is associated with alterations

in reproductive tract structure and function, including occurrence of uterine fibroids (D'Aloisio et

al. 2012), endometriosis (Upson et al. 2015), and early age at menarche (Adgent et al. 2012). In

rodents, perinatal genistein exposure can cause a variety of subsequent adverse consequences to

the reproductive system including altered estrous cycle, subfertility/infertility, delayed vaginal

opening, ovarian dysfunction, and uterine adenocarcinoma (Jefferson et al. 2002; Newbold et al. 2001; Patisaul et al. 2014), along with systemic effects including development of obesity (Strakovsky et al. 2014). Many of these effects can be observed at serum concentrations similar to those experienced by infants fed soy formula (Cao et al. 2009). Other xenoestrogens can produce delayed health effects in humans, with the best-studied example being prenatal diethylstilbestrol (DES) exposure and vaginal cancer in adolescence (Hoover et al. 2011). Although the mechanism by which soy formula produces late reproductive effects in humans remains unknown, epigenetic changes have been proposed as a mechanism by which DES acts (Hilakivi-Clarke et al. 2013). Indeed, neonatal mouse exposure to DES induces epigenetic changes in the uterus that are persistently maintained and are associated with altered gene expression in adults (Jefferson et al. 2013). In addition, prenatal exposure to genistein leads to gene silencing via DNA methylation in the agouti mouse model (Dolinoy et al. 2006), along with permanent downregulation of estrogen-responsive genes and hypermethylation of repetitive elements in prenatally-exposed adult animals (Vanhees et al. 2011). Postnatal genistein exposure is associated with hypermethylation and decreased gene expression in relation to obesity in a murine model (Strakovsky et al. 2014). Epigenetic changes are also a prominent feature of cancer cells, and phytoestrogens are receiving increasing attention, including clinical trials, as epigenetic re-programmers for cancer prevention and treatment in adults (Greenwald 2004; Pudenz et al. 2014). But what might be positive attributes in the setting of adult cancer may have different consequences in pre- and postnatal life where epigenetic landscapes are rapidly shifting (Gluckman et al. 2008). Phytoestrogens affect epigenetic programs in differentiating embryonic stem cells (Sato et al. 2011), developing embryos (Chan 2009; Dolinoy 2008), and as we observed here, infants who consume soy formula.

According to the American Academy of Pediatrics, soy formula is specifically indicated for

galactosemia and lactase deficiency (Bhatia et al. 2008). Families whose children do not have

these conditions use it for a variety of reasons, such as successful use of soy-formula products

with a previous child, dietary practices, or family recommendations (Stang et al. 2010). In

addition, soy formula is often a second formula given to children several months old after

cessation of breast feeding or in response to new onset of gastrointestinal symptoms. These

children may receive lower doses or be less susceptible to epigenetic reprograming than the

infant girls included in this study. While the epigenetic data presented here will augment this

discussion, we do not believe they constitute a clear contraindication for soy formula use.

Our study was limited to a single tissue from girls, examined only a small proportion of the

approximately 28 million CpG sites in the genome, and did not examine non-CpG methylation or

other epigenetic modifications. Consequently, the associations we find between soy formula and

increased methylation at *PRR5L* should be viewed as an initial and exploratory example of such

effects rather than a complete catalogue. Rapid advances in DNA and RNA sequencing

technologies can provide more comprehensive coverage of the epigenome and effects on gene

expression, but such assays were not possible in the current study because of the limited number

of cells available for analysis.

In summary, we found epigenetic differences in serial samples of vaginal cells from soy-

formula-fed compared to cow-formula-fed infant girls. This finding is consistent with the ability

of genistein, the principal phytoestrogen in soy formula, to act as an estrogen and to produce

epigenetic alterations in animal models. The search for epigenetic effects could be extended to

additional regions of the genome, other putative endocrine disruptors, different points in life

following exposure, and different tissues. Our results provide additional support for the hypothesis that epigenetic modification may be a mechanism by which early life exposures lead to later life health effects.

REFERENCES

Adgent M, Umbach D, Zemel B, Kelly A, Ford E, Bernbaum J, et al. 2014. Soy-based infant feeding is associated with estrogenized urogenital epithelium in girls at 24 weeks of age. In: Abstracts of the 2014 Conference of the International Society of Environmental Epidemiology (ISEE). Abstract 2578. Research Triangle Park, NC:Environmental Health Perspectives; http://dx.doi.org/10.1289/ehp.isee2014.

Adgent MA, Daniels JL, Rogan WJ, Adair L, Edwards LJ, Westreich D, et al. 2012. Early-life soy exposure and age at menarche. Paediatr Perinat Epidemiol 26:163-175.

Adgent MA, Flake GP, Umbach DM, Stallings VA, Bernbaum JC, Rogan WJ. 2013. Urogenital epithelial cells as simple markers of estrogen response in infants: Methods and applications. PLoS One 8:e77061.

Ahmad A, Biersack B, Li Y, Kong D, Bao B, Schobert R, et al. 2013. Deregulation of pi3k/akt/mtor signaling pathways by isoflavones and its implication in cancer treatment. Anticancer Agents Med Chem 13:1014-1024.

Andres A, Cleves MA, Bellando JB, Pivik RT, Casey PH, Badger TM. 2012. Developmental status of 1-year-old infants fed breast milk, cow's milk formula, or soy formula. Pediatrics 129:1134-1140.

Andres A, Moore MB, Linam LE, Casey PH, Cleves MA, Badger TM. 2015. Compared with feeding infants breast milk or cow-milk formula, soy formula feeding does not affect subsequent reproductive organ size at 5 years of age. J Nutr 145:871-875.

Behr M, Oehlmann J, Wagner M. 2011. Estrogens in the daily diet: In vitro analysis indicates that estrogenic activity is omnipresent in foodstuff and infant formula. Food Chem Toxicol 49:2681-2688.

Bergman A, Heindel JJ, Kasten T, Kidd KA, Jobling S, Neira M, et al. 2013. The impact of endocrine disruption: A consensus statement on the state of the science. Environ Health Perspect 121:A104-106.

Bernbaum JC, Umbach DM, Ragan NB, Ballard JL, Archer JI, Schmidt-Davis H, et al. 2008. Pilot studies of estrogen-related physical findings in infants. Environ Health Perspect 116:416-420.

Bhatia J, Greer F, American Academy of Pediatrics Committee on N. 2008. Use of soy protein-based formulas in infant feeding. Pediatrics 121:1062-1068.

Bidlingmaier F, Wagner-Barnack M, Butenandt O, Knorr D. 1973. Plasma estrogens in childhood and puberty under physiologic and pathologic conditions. Pediatr Res 7:901-907.

Broad Institute TCGA Genome Data Analysis Center. 2015. Correlation between mRNA

expression and DNA methylation. Broad Institute of MIT and Harvard.

doi:10.7908/C1RV0MWM, 10.7908/C1WM1CQP [accessed 18 March 2016].

Cao Y, Calafat AM, Doerge DR, Umbach DM, Bernbaum JC, Twaddle NC, et al. 2009.

Isoflavones in urine, saliva, and blood of infants: Data from a pilot study on the estrogenic

activity of soy formula. J Expo Sci Environ Epidemiol 19:223-234.

Carrick DM, Lai WS, Blackshear PJ. 2004. The tandem ccch zinc finger protein tristetraprolin

and its relevance to cytokine mrna turnover and arthritis. Arthritis Res Ther 6:248-264.

Chan WH. 2009. Impact of genistein on maturation of mouse oocytes, fertilization, and fetal

development. Reprod Toxicol 28:52-58.

Chiaroni-Clarke RC, Munro JE, Chavez RA, Pezic A, Allen RC, Akikusa JD, et al. 2014.

Independent confirmation of juvenile idiopathic arthritis genetic risk loci previously identified by

immunochip array analysis. Pediatr Rheumatol Online J 12:53.

D'Aloisio AA, Baird DD, DeRoo LA, Sandler DP. 2012. Early-life exposures and early-onset

uterine leiomyomata in black women in the sister study. Environ Health Perspect 120:406-412.

D'Aloisio AA, DeRoo LA, Baird DD, Weinberg CR, Sandler DP. 2013. Prenatal and infant

exposures and age at menarche. Epidemiology 24:277-284.

Day JK, Bauer AM, DesBordes C, Zhuang Y, Kim BE, Newton LG, et al. 2002. Genistein alters

methylation patterns in mice. J Nutr 132:2419S-2423S.

Diamanti-Kandarakis E, Bourguignon JP, Giudice LC, Hauser R, Prins GS, Soto AM, et al. 2009. Endocrine-disrupting chemicals: An endocrine society scientific statement. Endocr Rev 30:293-342.

Doerge DR, Twaddle NC, Banks EP, Jefferson WN, Newbold RR. 2002. Pharmacokinetic analysis in serum of genistein administered subcutaneously to neonatal mice. Cancer Lett 184:21-27.

Dolinoy DC, Weidman JR, Waterland RA, Jirtle RL. 2006. Maternal genistein alters coat color and protects avy mouse offspring from obesity by modifying the fetal epigenome. Environ Health Perspect 114:567-572.

Dolinoy DC. 2008. The agouti mouse model: An epigenetic biosensor for nutritional and environmental alterations on the fetal epigenome. Nutr Rev 66 Suppl 1:S7-11.

Gan X, Wang J, Wang C, Sommer E, Kozasa T, Srinivasula S, et al. 2012. Prr5l degradation promotes mtorc2-mediated pkc-delta phosphorylation and cell migration downstream of galpha12. Nat Cell Biol 14:686-696.

Gluckman PD, Hanson MA, Cooper C, Thornburg KL. 2008. Effect of in utero and early-life conditions on adult health and disease. N Engl J Med 359:61-73.

Greenwald P. 2004. Clinical trials in cancer prevention: Current results and perspectives for the future. J Nutr 134:3507S-3512S.

Hewitt SC, Li L, Grimm SA, Chen Y, Liu L, Li Y, et al. 2012. Research resource: Whole-genome estrogen receptor alpha binding in mouse uterine tissue revealed by chip-seq. Mol Endocrinol 26:887-898.

Hilakivi-Clarke L, de Assis S, Warri A. 2013. Exposures to synthetic estrogens at different times during the life, and their effect on breast cancer risk. J Mammary Gland Biol Neoplasia 18:25-42.

Holmes B, Artinian N, Anderson L, Martin J, Masri J, Cloninger C, et al. 2012. Protor-2 interacts with tristetraprolin to regulate mrna stability during stress. Cell Signal 24:309-315.

Hoover RN, Hyer M, Pfeiffer RM, Adam E, Bond B, Cheville AL, et al. 2011. Adverse health outcomes in women exposed in utero to diethylstilbestrol. N Engl J Med 365:1304-1314.

Jefferson WN, Couse JF, Padilla-Banks E, Korach KS, Newbold RR. 2002. Neonatal exposure to genistein induces estrogen receptor (er)alpha expression and multioocyte follicles in the maturing mouse ovary: Evidence for erbeta-mediated and nonestrogenic actions. Biol Reprod 67:1285-1296.

Jefferson WN, Patisaul HB, Williams CJ. 2012. Reproductive consequences of developmental phytoestrogen exposure. Reproduction 143:247-260.

Jefferson WN, Chevalier DM, Phelps JY, Cantor AM, Padilla-Banks E, Newbold RR, et al. 2013. Persistently altered epigenetic marks in the mouse uterus after neonatal estrogen exposure. Mol Endocrinol 27:1666-1677.

Klein SL, Wisniewski AB, Marson AL, Glass GE, Gearhart JP. 2002. Early exposure to genistein exerts long-lasting effects on the endocrine and immune systems in rats. Mol Med 8:742-749.

Kuiper GG, Lemmen JG, Carlsson B, Corton JC, Safe SH, van der Saag PT, et al. 1998. Interaction of estrogenic chemicals and phytoestrogens with estrogen receptor beta. Endocrinology 139:4252-4263.

McCarver G, Bhatia J, Chambers C, Clarke R, Etzel R, Foster W, et al. 2011. Ntp-cerhr expert panel report on the developmental toxicity of soy infant formula. Birth Defects Res B Dev Reprod Toxicol 92:421-468.

Newbold RR, Banks EP, Bullock B, Jefferson WN. 2001. Uterine adenocarcinoma in mice treated neonatally with genistein. Cancer Res 61:4325-4328.

Patisaul HB, Mabrey N, Adewale HB, Sullivan AW. 2014. Soy but not bisphenol a (bpa) induces hallmarks of polycystic ovary syndrome (pcos) and related metabolic co-morbidities in rats. Reprod Toxicol 49C:209-218.

Pudenz M, Roth K, Gerhauser C. 2014. Impact of soy isoflavones on the epigenome in cancer prevention. Nutrients 6:4218-4272.

Riu A, Balaguer P, Perdu E, Pandelova M, Piccinelli R, Gustafsson JA, et al. 2008. Characterisation of bioactive compounds in infant formulas using immobilised recombinant estrogen receptor-alpha affinity columns. Food Chem Toxicol 46:3268-3278.

Rossen LM, Simon AE, Herrick KA. 2015. Types of infant formulas consumed in the united states. Clin Pediatr (Phila) pii: 0009922815591881 [Online 6 July 2015].

Sato N, Yamakawa N, Masuda M, Sudo K, Hatada I, Muramatsu M. 2011. Genome-wide DNA methylation analysis reveals phytoestrogen modification of promoter methylation patterns during embryonic stem cell differentiation. PLoS One 6:e19278.

Stang J, Hoss K, Story M. 2010. Health statements made in infant formula advertisements in pregnancy and early parenting magazines: A content analysis. ICAN: Infant Child Adolesc Nutr 2:16-25.

Strakovsky RS, Lezmi S, Flaws JA, Schantz SL, Pan YX, Helferich WG. 2014. Genistein exposure during the early postnatal period favors the development of obesity in female, but not male rats. Toxicol Sci 138:161-174.

Taylor GA, Carballo E, Lee DM, Lai WS, Thompson MJ, Patel DD, et al. 1996. A pathogenetic role for tnf alpha in the syndrome of cachexia, arthritis, and autoimmunity resulting from tristetraprolin (ttp) deficiency. Immunity 4:445-454.

Thedieck K, Polak P, Kim ML, Molle KD, Cohen A, Jeno P, et al. 2007. Pras40 and prr5-like protein are new mtor interactors that regulate apoptosis. PLoS One 2:e1217.

Upson K, Sathyanarayana S, Scholes D, Holt VL. 2015. Early-life factors and endometriosis risk. Fertil Steril 104:964-971.

Vanhees K, Coort S, Ruijters EJ, Godschalk RW, van Schooten FJ, Barjesteh van Waalwijk van Doorn-Khosrovani S. 2011. Epigenetics: Prenatal exposure to genistein leaves a permanent signature on the hematopoietic lineage. FASEB J 25:797-807.

Venables WN, Ripley BD. 2002. Modern applied statistics with s. New York:Springer

Wang X, Ning Y, Guo X. 2015. Integrative meta-analysis of differentially expressed genes in osteoarthritis using microarray technology Mol Med Rep 12:3439-3445.

TABLES

Table 1. Participant characteristics

Characteristic	Cow formula N (%)	Soy Formula N (%)	p-value*
Total	22 (100)	28 (100)	
Maternal Age			
20 or less	4 (18)	4 (14)	0.78
21-30	12 (55)	18(64)	
31+	6 (27)	6 (21)	
Child's Race			
Black	15 (68)	20 (71)	0.74
White	5 (23)	4 (14)	
Multiple/Other	2 (9)	4 (14)	
Mother's Race			
Black	17 (77)	20 (74)	0.79
White	5 (23)	7 (26)	
Missing (Unknown)	0	1	
Maternal Education			
High School /GED[†] or less	15 (68)	14 (50)	0.19
Some College or More	7 (32)	14 (50)	

*P-values are from chi-square tests (maternal age, maternal education) or Fisher's exact tests

(mother's race, child's race).

†General Educational Development

FIGURE LEGENDS

Figure 1A. Data from EWAS analysis. Volcano plot depicting all CpGs present on the 450K methylation array: log10 (p-value) plotted against effect size. Three CpGs in the gene *PRR5L* (circled) were identified as having small p-values in combination with large effect sizes.

Figure 1B. Data from EWAS analysis. The 5' region of *PRR5L* transcript variant 4 (Chr 11) showing CpGs and methylation levels obtained from 450K arrays and the three CpGs with largest effect sizes and smallest p values, marked with*, from the volcano plot. CpGs validated by pyrosequencing are marked with P. DNA methylation levels were significantly higher in vaginal DNA samples from soy-formula-fed girls.

Figure 1C. Initial replication of 450K array results by pyrosequencing. All 11 individual samples that yielded usable data from the 450K array were re-run using pyrosequencing specific for cg00220721 and cg2211705. One sample (cow-formula-fed) failed in the pyrosequencing run due to low input amount. The other 10 showed strong correlation (R^2=0.96) between the average the two CpGs from array versus pyrosequencing measurements.

Figure 2A. Data from Pyrosequencing. Replication of the top two *PRR5L* CpG sites by pyrosequencing. Each dot represents the average of the two adjacent CpG methylation values from two separate pyrosequencing runs for each sample (red dots=soy-formula-fed girls, blue dots=cow-formula-fed girls). Lines represent the fitted mean trajectory for the corresponding feeding group; shading represents 95% pointwise confidence intervals. Estimated mean trajectories are natural cubic splines with three knots fitted to the data via mixed-model techniques.

Figure 2B. Data from Pyrosequencing. The difference between the fitted mean methylation at individual time points. Black line represents the difference between soy and cow formula values calculated from Figure 2A at each time point. Thin blue lines represent 95% pointwise confidence intervals for each time point.

Figure 3. Relative expression of mouse *Prr5l*. Mean expression in untreated newborn mice was used as the reference level. Error bars represent standard error of the mean. Mice were treated with daily subcutaneous injections of genistein (50 mg/kg) or vehicle control on postnatal days 1-5.
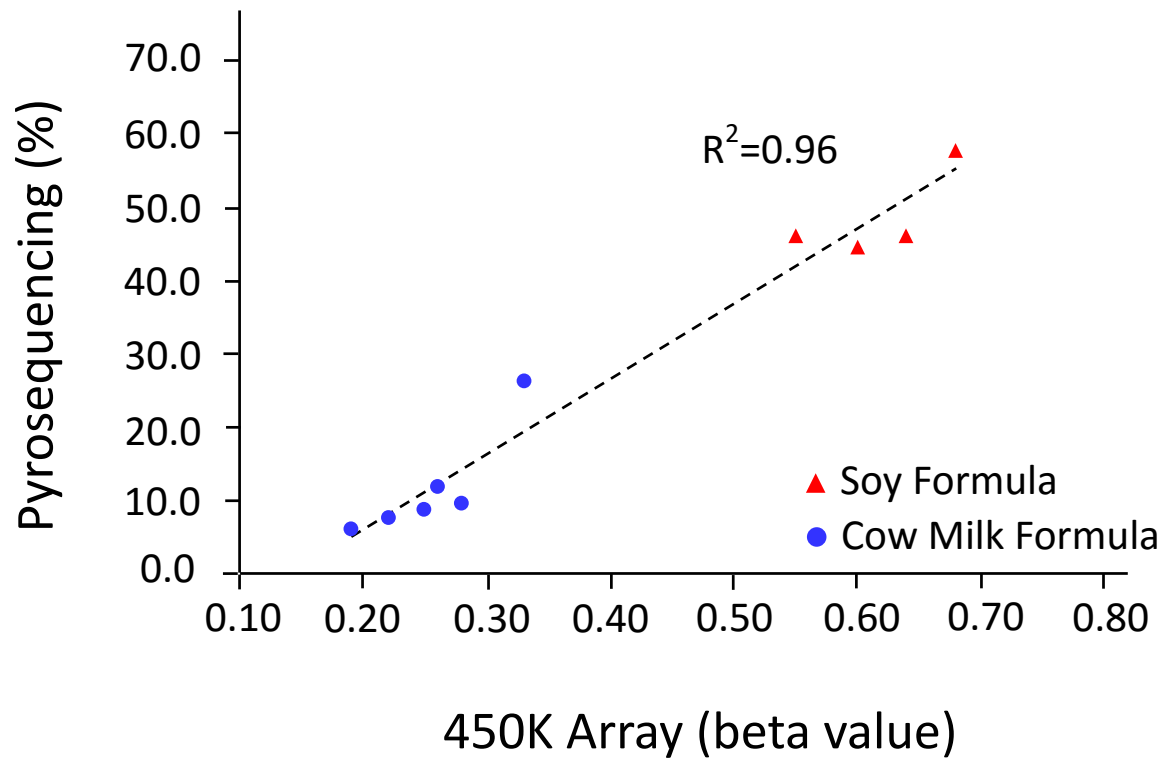
Figure 1a.
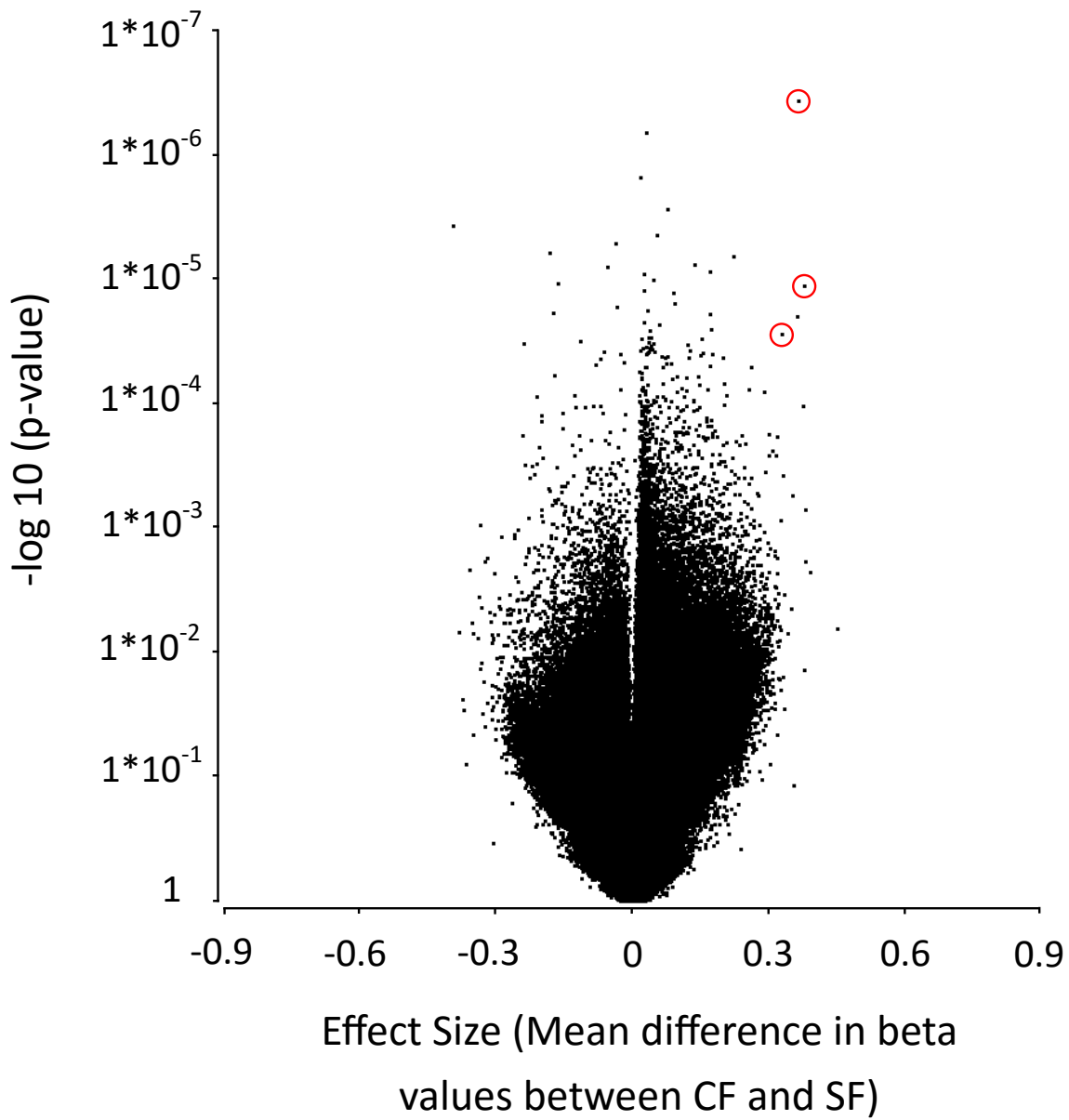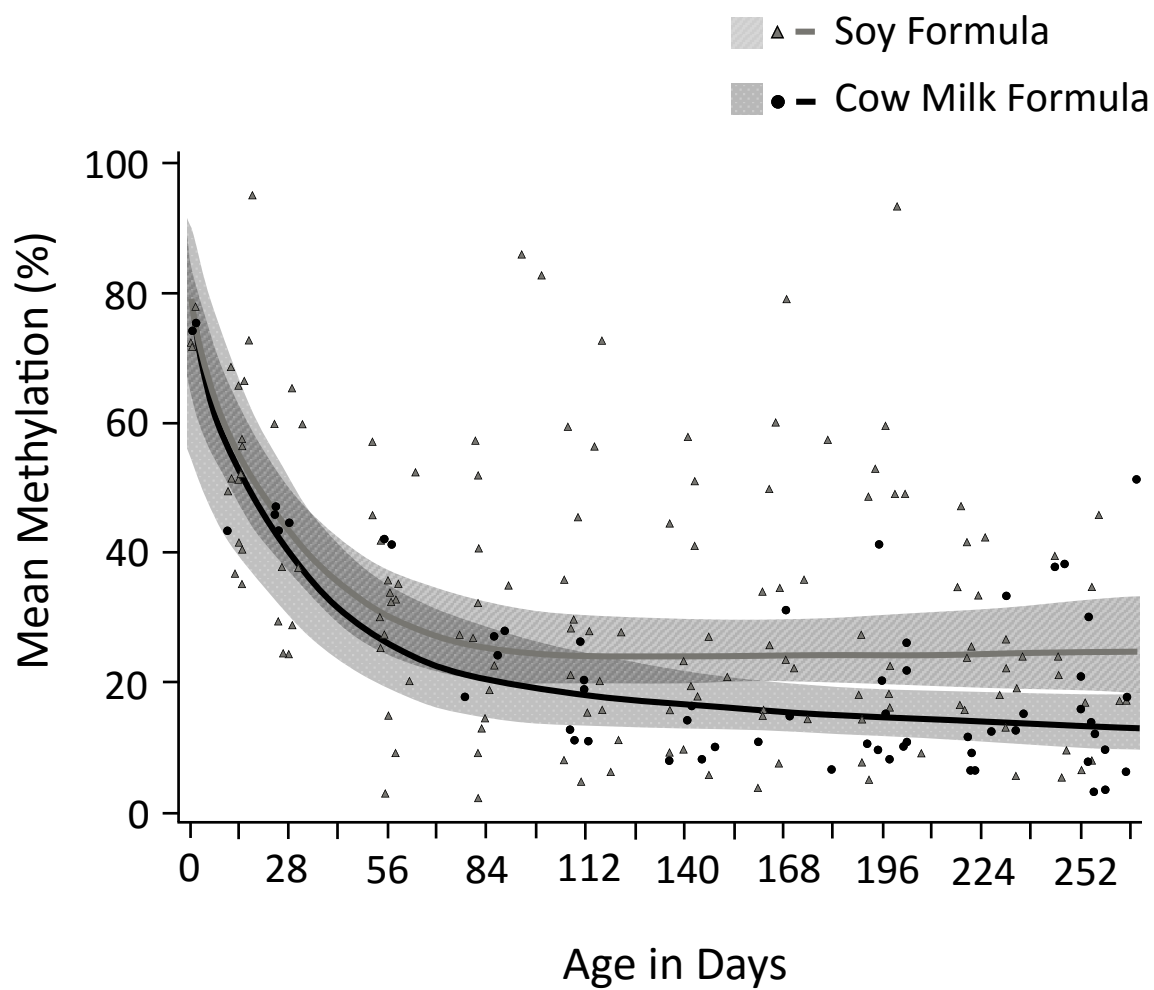
Figure 1b.
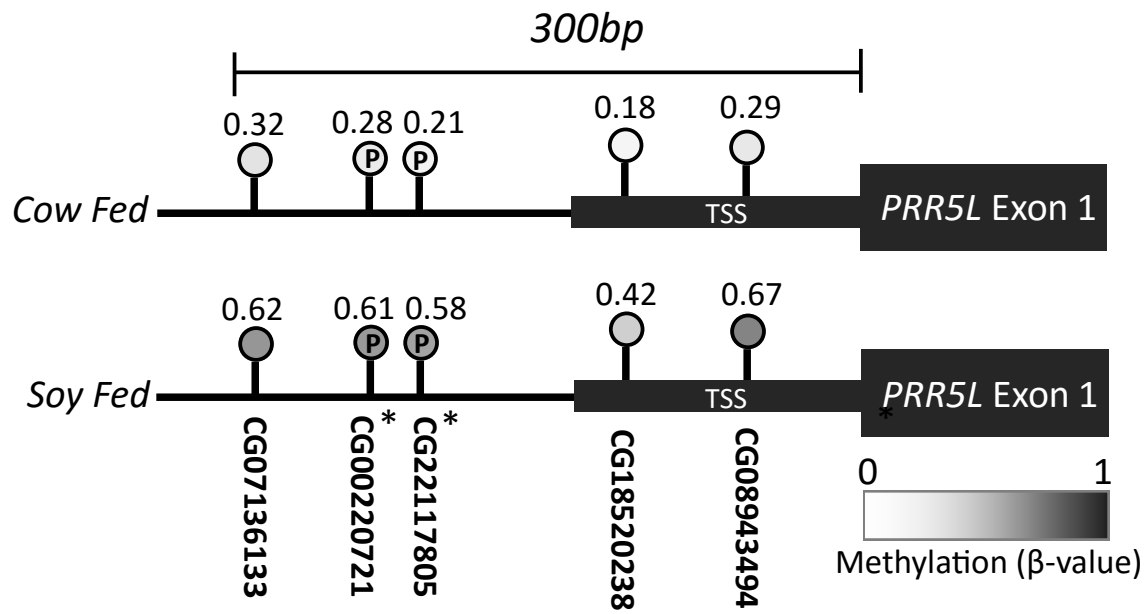
Figure 1c.

Figure 2a.

Figure 2b.

Figure 3.